CLAIMS

What is claimed is:

1. A switch comprising:

a memory configured to store connection information;

a server address translator configured to receive a plurality of requests from a client over a single connection, to reference the memory to determine a plurality of servers to service said received plurality of requests; and to redirect said received plurality of requests to said determined plurality of servers; and

a client address translator configured to receive a plurality of responses from said determined plurality of servers; to organize said received plurality of responses into a stream of packets; and to forward said stream of packets over the connection to the client.

2. The switch of claim 1, wherein the server address translator is further configured to send a plurality of splicer tokens to said determined plurality of servers; and wherein the client address translator is further configured to receive a plurality of splicer token responses; and to update the memory in response to said receipt of the plurality of splicer token responses.

3. A method comprising:

receiving a plurality of requests from a client over a single Transmission Control Protocol (TCP) connection;

redirecting the plurality of requests to a plurality of servers;

receiving a plurality of responses from the plurality of servers;

organizing the plurality of the responses into a stream of packets; and

sending the stream of packets to the client over the single connection.

4. The method of claim 3, further comprising:

sending a plurality of splicer indications to the plurality of servers; and

receiving a plurality of splicer responses from the plurality of servers; and

wherein said organizing the plurality of the responses includes referencing the plurality of splicer responses.

5. The method of claim 4, further comprising updating the memory or a second memory in response to receiving the plurality of splicer responses.

6. A computer-readable medium containing computer-executable instructions for performing the method of claim 3.

7. A packet switch performing the method of claim 3.

8. A router performing the method of claim 3.

9. A method comprising:

receiving a first request over a connection from a client;

redirecting the first request to a first server;

receiving a first response to the first request from the first server;

forwarding the first response over the connection to the client;

receiving a second request over the connection from the client before said forwarding the first response;

redirecting the second request to a second server;

receiving a second response to the second request from the second server; and

forwarding the second response over the connection to the client.

10. The method of claim 9, further comprising:

sending a first splicer token to the first server after redirecting the first request to the first server; and

receiving a first splicer token response from the first server.

11. The method of claim 10, further comprising updating a memory for storing splicer data after receiving the first splicer response.

12. The method of claim 11, wherein the splicer data indicates an address of the client.

13. The method of claim 11, wherein the splicer data indicates a sequence number for a set of packets received from the client.

14. The method of claim 9, further comprising selecting the first server from a set of server identifiers maintained in a memory configured to store connection information.

15. A computer-readable medium containing computer-executable instructions for performing the method of claim 9.

16. A method comprising:

establishing a set of connections to a plurality of servers;

maintaining an indication of the set of connections;

receiving a first request over a Transmission Control Protocol (TCP) connection from a client;

referencing the indication to determine a first one of the plurality of servers;

redirecting the first request to the first one of the plurality of servers;

receiving a first response to the first request from the first one of the plurality of servers;

receiving a second request over the connection from the client before said receiving the first response;

referencing the indication to determine a second one of the plurality of servers;

redirecting the second request to the second one of the plurality of servers;

receiving a second response to the second request from the second one of the plurality of servers; and

organizing the first and second responses into a stream of packets.

17. The method of claim 16, further comprising:

sending a splicer token to the first one of the plurality of servers after redirecting the first request to the first one of the plurality of servers;

receiving a splicer token response from the first one of the plurality of servers; and

updating the indication in response to receiving the splicer token response.

18. A computer-readable medium containing computer-executable instructions for performing the method of claim 16.

19. An apparatus comprising:

means for receiving a plurality of requests from a client over a single connection;

means for redirecting the plurality of requests to a plurality of servers;

means for receiving a plurality of responses from the plurality of servers;

means for organizing the plurality of the responses into a stream of packets; and

means for sending the stream of packets to the client over the single connection.

20. The apparatus of claim 19, further comprising:

means for sending a plurality of splicer indications to the plurality of servers; and

means for receiving a plurality of splicer responses from the plurality of servers; and

wherein said means for organizing the plurality of the responses includes means for referencing the plurality of splicer responses.

21. An apparatus comprising:

means for establishing a set of connections to a plurality of servers;

means for maintaining a data structure indicating the set of connections;

means for receiving a first request over a connection from a client;

means for referencing the data structure to determine a first one of the plurality of servers;

means for redirecting the first request to the first one of the plurality of servers;

means for receiving a first response to the first request from the first one of the plurality of servers;

means for receiving a second request over the connection from the client before said receiving the first response;

means for referencing the data structure to determine a second one of the plurality of servers;

means for redirecting the second request to the second one of the plurality of servers;

means for receiving a second response to the second request from the second one of the plurality of servers;

means for organizing the first and second responses into a stream of packets; and

means for forwarding the stream of packets to the client.